

# TITLE OF THE INVENTION

NONVOLATILE SEMICONDUCTOR MEMORY CAPABLE OF GENERATING  
READ-MODE REFERENCE CURRENT AND VERIFY-MODE REFERENCE  
CURRENT FROM THE SAME REFERENCE CELL

## 5 CROSS-REFERENCE TO RELATED APPLICATIONS

This application is based upon and claims the  
benefit of priority from the prior Japanese Patent  
Applications No. 2002-206177, filed July 15, 2002; and  
No. 2003-193728, filed July 8, 2003, the entire  
10 contents of both of which are incorporated herein by  
reference.

## BACKGROUND OF THE INVENTION

### 1. Field of the Invention

The present invention relates to a nonvolatile  
15 semiconductor memory. More specifically, the invention  
relates to a multilevel flash memory wherein a multi-  
level of three or more levels is stored in a single  
cell.

### 2. Description of the Related Art

20 A binary flash memory for storing two data  
items of different levels (referred to as binary  
data hereinafter) widely spreads as a nonvolatile  
semiconductor memory.

FIG. 16 shows the arrangement of the main part of  
25 a binary flash memory (e.g., a NOR type). Referring to  
FIG. 16, a cell array 101 includes a plurality of  
memory cells (main cells) MC arranged in matrix.

The control gates of memory cells MC arranged in one row are connected to a common one of a plurality of word lines WL0 to WLn. The drain regions of memory cells MC arranged in one column are connected to a common one of a plurality of bit lines BL0 to BLk. Generally, the cell array 101 is divided into a plurality of blocks. The source regions of memory cells MC in one block are connected to a common one of a plurality of source lines (not shown). The bit lines BL0 to BLk are connected to a sense amplifier 102 through their corresponding one of a plurality of select transistors ST0 to STk. A plurality of column lines COL0 to COLm are connected to the gates of the select transistors ST0 to STk, respectively.

A reference circuit 103 includes one reference cell RC and a plurality of dummy cells DC. The drain regions of the reference cell RC and dummy cells DC are connected to each other. The control gate of the reference cell RC is connected to a reference word line RWL. The drain region of the reference cell RC is connected to the sense amplifier 102 via an n-type MOS transistor 103a. A reference column line RCOL is connected to the gate of the transistor 103a.

The sense amplifier 102 includes n-type MOS transistors 102a and 102b, p-type MOS transistors 102c and 102d and a differential amplifier 102e. The source region of the n-type MOS transistor 102a is connected

to the drain regions of the select transistors ST0 to STk in the cell array 101. The drain region of the n-type MOS transistor 102a is connected to the gate and the drain region of the p-type MOS transistor 102c and the inverted input terminal of the differential amplifier 102e. On the other hand, the source region of the n-type MOS transistor 102b is connected to the drain region of the n-type MOS transistor 103a in the reference circuit 103. The drain region of the n-type MOS transistor 102b is connected to the gate and the drain region of the p-type MOS transistor 102d and the noninverted input terminal of the differential amplifier 102e. The differential amplifier 102e outputs sensed cell data (Dout) from its output terminal.

A BIAS power supply voltage is applied to the gate of each of the n-type MOS transistors 102a and 102b. A power supply voltage Vcc is applied to the source region of each of the p-type MOS transistors 102c and 102d.

FIG. 17 shows the characteristics of the binary flash memory with the above arrangement. More specifically, FIG. 12 shows a relationship ( $V_g$ - $I_{cell}$  ( $I_d$ ) characteristics) between a gate voltage  $V_g$  applied to the control gate and a cell current  $I_{cell}$  (drain current  $I_d$ ) per load current in both the memory cell MC and the reference cell RC. In data read and program

verify modes, a read voltage  $V_r$  is applied to the control gate of the reference cell RC. Thus, the sense amplifier 102 always determines the cell current  $I_{cell}$  ("0" or "1") by the reference current  $I_{ref}$ .

5           In the  $V_g$ - $I_d$  characteristics of the memory cell MC, the state of a relatively large number of electrons stored in a floating gate (or a high threshold voltage  $V_{th}$  of the memory cell MC) is considered to be data "0". The memory cell MC that stores the data "0" is referred to as a "0" cell. Conversely, the state of  
10           a relatively small number of electrons (or a low threshold voltage  $V_{th}$  of the memory cell MC) is considered to be data "1". The memory cell MC that stores the data "1" is referred to as a "1" cell.

15           The cell current of the reference cell RC (reference current  $I_{ref}$ ) is set to approximately half the cell current  $I_{cell}$  of the memory cell MC. In other words, when the gate voltage  $V_g$  is equal to the read voltage  $V_r$ , a difference between the cell current  $I_{cell}$   
20           of the memory cell MC and the cell current  $I_{ref}$  of the reference cell RC in the "0" cell and that in the "1" cell are almost equal to each other.

          FIG. 18 shows a correlation between the gate voltage ( $V_g$ -hontai) of the memory cell MC and the gate  
25           voltage ( $V_g$ -ref) of the reference cell RC in each of operating modes. For example, in program verify (PV) mode for defining a data program state "0", a program

verify voltage  $V_{pv}$  ( $= 6.5V$ ) is applied to the control gate of the memory cell MC. A difference between the program verify voltage  $V_{pv}$  and the read voltage  $V_r$  ( $= 5.5V$ ) that is applied to the control gate of the reference cell RC is reflected in the threshold voltage  $V_{th}$  of the memory cell MC. Thus, the "0" cell is usually cut off when the gate voltage  $V_g$  is equal to the read voltage  $V_r$ . Similarly, for example, in erase verify mode (EV) mode for defining a data erase state "1", an erase verify voltage  $V_{ev}$  ( $= 4V$ ) is applied to the control gate of the memory cell MC. If the erase verify voltage  $V_{ev}$  is set at roughly the same as the reference voltage  $V_{tref}$ , the current flowing when the gate voltage  $V_g$  of the "1" cell is equal to the read voltage  $V_r$ , i.e., the cell current  $I_{cell}$  becomes almost equal to  $2I_{ref}$ .

In other words, the sense amplifier 102 senses a current difference ( $+I_{ref}/-I_{ref}$ ) between the cell current  $I_{cell}$  of the memory cell MC and the cell current  $I_{ref}$  of the reference cell RC, which is caused when the gate voltage  $V_g$  is equal to the read voltage  $V_r$  and converts it into a digital signal of "0" or "1". Accordingly, cell data is read out.

FIG. 19 shows the distribution of threshold voltages  $V_{th}$  with respect to the gate voltage  $V_g$  in the memory cell (binary cell) MC capable of storing binary data "0" and "1". For example, in over-erase verify

(OEV) mode for compensating for the lower limit of the cell distribution corresponding to the data erase state "1", an over-erase verify voltage  $V_{oev}$  ( $= 2V$ ), which is lower than the erase verify voltage  $V_{ev}$ , is applied to the control gate of the memory cell MC. A cell that becomes "1" when the over-erase verify voltage  $V_{oev}$  is applied is detected. Data is written such that the threshold voltage  $V_{th}$  of a bit corresponding to the cell becomes "0" when the over-erase verify voltage  $V_{oev}$  is applied. Thus, the cell distribution corresponding to the erase state "1" falls within a given range.

The cell area per bit is a cost index of a flash memory. There is a multilevel flash memory that can be reduced in cost by storing data of a plurality of bits in one cell as well as by decreasing the cell area. A four-level flash memory for storing four levels or four data items of different levels (four-level data) has been already reported in, for example, M. bauer et al., "A Multilevel-Cell 32Mb Flash Memory", ISSCC Digest of Technical Papers, pp. 132-133, 1995.

The cell distribution of the four-level flash memory is discrete with respect to the gate voltage. More specifically, in verify mode, a margin for reading is secured by varying the gate voltage of a memory cell as in the binary flash memory. On the other hand, cell data is read out at a constant gate voltage.

Assuming that the memory cells vary in transconductance or the ratio of a change in cell current to that in gate voltage varies, a margin (read margin) for the reference current of read current will be lowered. The low read margin influences the read access time and causes a failure in reading.

As described above, the prior art multilevel flash memory secures a verify margin at the gate voltage. For this reason, there occurs a problem that the variations in the ratio of a change in cell current to that in gate voltage lowers the read margin.

#### BRIEF SUMMARY OF THE INVENTION

According to a first aspect of the present invention, there is provided a nonvolatile semiconductor memory comprising: a plurality of nonvolatile memory cells each having a gate, a drain and a source to hold data corresponding to a threshold voltage level; a plurality of word lines connected to gates of the nonvolatile memory cells, respectively; a plurality of bit lines connected to drains of the nonvolatile memory cells, respectively; a plurality of source lines connected to sources of the nonvolatile memory cells, respectively; a reference current generation circuit which generates a reference current, the reference current generation circuit including at least one reference cell and an amplification circuit which amplifies a current flowing through the reference cell,

and a ratio of an amplification factor of current in  
a program verify mode to an amplification factor of  
current in a data read mode is larger than 1; and  
a sense amplifier which compares the reference current  
5 with a current flowing through selected ones of the  
nonvolatile memory cells and reads data held in the  
selected ones of the nonvolatile memory cells.

According to a second aspect of the present  
invention, there is provided a nonvolatile semicon-  
10 ductor memory comprising: a plurality of nonvolatile  
memory cells each having a gate, a drain and a source  
to hold a multilevel of three or more levels  
corresponding to a threshold voltage level; a plurality  
of word lines connected to gates of the nonvolatile  
15 memory cells, respectively; a plurality of bit lines  
connected to drains of the nonvolatile memory cells,  
respectively; a plurality of source lines connected to  
sources of the nonvolatile memory cells, respectively;  
a reference current generation circuit which generates  
20 at least first and second reference currents, the  
reference current generation circuit including at least  
a first reference cell, a second reference cell having  
a threshold voltage that is higher than that of the  
first reference cell, a first amplification circuit  
25 which amplifies a current flowing through the first  
reference cell, and a second amplification circuit  
which amplifies a current flowing through the second



reference cell, a first current amplification ratio of  
an amplification factor of current in program verify  
mode to an amplification factor of current in a data  
read mode in the first amplification circuit being  
5 larger than 1, a second current amplification ratio of  
an amplification factor of current in a program verify  
mode to an amplification factor of current in a data  
read mode in the second amplification circuit being  
larger than 1, and the first current amplification  
10 ratio being smaller than the second current amplifica-  
tion ratio; a first sense amplifier which compares the  
first reference current with a current flowing through  
a selected one of the nonvolatile memory cells and  
reads a signal corresponding to a multilevel held in  
15 the selected one of the nonvolatile memory cells; and  
a second sense amplifier which compares the second  
reference current with a current flowing through  
a selected one of the nonvolatile memory cells and  
reads a signal corresponding to a multilevel held in  
20 the selected one of the nonvolatile memory cells.

According to a third aspect of the present  
invention, there is provided a nonvolatile semicon-  
ductor memory comprising: a plurality of nonvolatile  
memory cells each having a gate, a drain and a source  
25 to hold a multilevel of three or more levels  
corresponding to a threshold voltage level; a plurality  
of word lines connected to gates of the nonvolatile

memory cells, respectively; a plurality of bit lines  
connected to drains of the nonvolatile memory cells,  
respectively; a plurality of source lines connected to  
sources of the nonvolatile memory cells, respectively;  
5 a reference current generation circuit which  
selectively generates at least first and second  
reference currents, the reference current generation  
circuit including at least a first reference cell,  
a second reference cell having a threshold voltage  
10 that is higher than that of the first reference cell,  
a first amplification circuit which amplifies a current  
flowing through the first reference cell, and a second  
amplification circuit which amplifies a current flowing  
through the second reference cell, a first current  
15 amplification ratio of an amplification factor of  
current in program verify mode to an amplification  
factor of current in a data read mode in the first  
amplification circuit is larger than 1, a second  
current amplification ratio of an amplification factor  
20 of current in a program verify mode to an amplification  
factor of current in a data read mode in the second  
amplification circuit is larger than 1, and the first  
current amplification ratio is smaller than the second  
current amplification ratio; and a sense amplifier  
25 which compares an output current of the reference  
current generation circuit and a cell current flowing  
through a selected one of the nonvolatile memory cells

and amplifies and outputs the cell current.

According to a fourth aspect of the present invention, there is provided a nonvolatile semiconductor memory comprising: a plurality of  
5 nonvolatile memory cells each having a gate, a drain and a source to hold a multilevel of three or more levels corresponding to a threshold voltage level; a plurality of word lines connected to gates of the nonvolatile memory cells, respectively; a plurality of  
10 bit lines connected to drains of the nonvolatile memory cells, respectively; a plurality of source lines connected to sources of the nonvolatile memory cells, respectively; a reference current generation circuit which selectively generates at least first and second  
15 reference currents, the reference current generation circuit including at least a first reference cell, a second reference cell having a threshold voltage that is higher than that of the first reference cell, a first amplification circuit which amplifies a current  
20 flowing through the first reference cell, and a second amplification circuit which amplifies a current flowing through the second reference cell, a first current amplification ratio of an amplification factor of current in program verify mode to an amplification  
25 factor of current in a data read mode in the first amplification circuit being larger than 1, a second current amplification ratio of an amplification factor

of current in a program verify mode to an amplification factor of current in a data read mode in the second amplification circuit being larger than 1, and the first current amplification ratio is smaller than the second current amplification ratio, the reference current generation circuit further including a third reference cell having a threshold voltage which is higher than that of the second reference cell in order to generate a third reference current, and a third amplification circuit which amplifies a current flowing through the third reference cell; and a sense amplifier which compares an output current of the reference current generation circuit and a cell current flowing through a selected one of the nonvolatile memory cells and amplifies and outputs the cell current, wherein the reference current generation circuit selectively outputs the second and third reference currents in accordance with logic of an output of the sense amplifier when the output current of the reference current generation circuit is the first reference current.

According to a fifth aspect of the present invention, there is provided a nonvolatile semiconductor memory comprising: a plurality of memory cells each having  $2^N$  (N is two or more) levels; a plurality of word lines connected to gates of the memory cells, respectively; a plurality of bit lines connected to

drains of the memory cells, respectively; a plurality of source lines connected to sources of the memory cells, respectively; a reference current generation circuit which selectively outputs one of (N-1) reference currents, the reference current generation circuit including (N-1) reference cells and (N-1) amplification circuits which amplify a current flowing through the (N-1) reference cells, a threshold voltage of a first reference cell of the (N-1) reference cells being higher than that of a (I-1)-th reference cell ( $1 \leq I \leq N$ ), a ratio of an I-th amplification factor of current in program verify mode to an amplification factor of current in a data read mode in an I-th amplification circuit of the (N-1) amplification circuits being larger than 1, and a (I-1)-th amplification factor being smaller than the I-th amplification factor; and a sense amplifier which compares an output current of the reference current generation circuit and a cell current flowing through a selected one of the memory cells and amplifies and outputs the cell current, wherein the reference current generation circuit selectively outputs the second and third reference currents in accordance with logic of an output of the sense amplifier when the output current of the reference current generation circuit is the first reference current.

According to a sixth aspect of the present

invention, there is provided a nonvolatile semiconductor memory comprising: a plurality of memory cells each having  $2^N$  ( $N$  is two or more) levels; a plurality of word lines connected to gates of the memory cells, respectively; a plurality of bit lines connected to drains of the memory cells, respectively; a plurality of source lines connected to sources of the memory cells, respectively; a reference current generation circuit which selectively outputs one of ( $N-1$ ) reference currents, the reference current generation circuit including ( $N-1$ ) reference cells and ( $N-1$ ) amplification circuits which amplify a current flowing through the ( $N-1$ ) reference cells, a threshold voltage of a first reference cell of the ( $N-1$ ) reference cells being higher than that of a ( $I-1$ )-th reference cell ( $1 \leq I \leq N$ ), a ratio of an  $I$ -th amplification factor of current in program verify mode to an amplification factor of current in a data read mode in an  $I$ -th amplification circuit of the ( $N-1$ ) amplification circuits being larger than 1, and a ( $I-1$ )-th amplification factor being smaller than the  $I$ -th amplification factor; and a sense amplifier which compares an output current of the reference current generation circuit and a cell current flowing through a selected one of the memory cells and amplifies and outputs the cell current, wherein the reference current generation circuit selectively outputs the second and

third reference currents in accordance with logic of an output of the sense amplifier when the output current of the reference current generation circuit is the first reference current.

5 BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING

FIG. 1 is a block diagram showing an example of an arrangement of a four-level flash memory according to an embodiment of the present invention;

10 FIG. 2 is a schematic sectional view of a structure of a memory cell in the four-level flash memory shown in FIG. 1;

FIG. 3 is a circuit diagram showing an arrangement of the main part (a sense amplifier and a reference current generation circuit) of the four-level flash  
15 memory shown in FIG. 1;

FIGS. 4A to 4C are circuit diagrams showing examples of arrangements of voltage generation circuits in the reference current generation circuit shown in FIG. 3;

20 FIG. 5 is a graph showing  $V_g$ - $I_{cell}$  characteristics of the four-level flash memory when the voltage generation circuits are arranged as shown in FIGS. 4A to 4C;

FIG. 6 is a diagram showing an example of the cell  
25 distribution with respect to a reference current in the four-level flash memory shown in FIG. 1;

FIG. 7 is a diagram showing a correlation between

a reference current and a gate voltage in each of operating modes when the voltage generation circuits are arranged as shown in FIGS. 4A to 4C;

FIGS. 8A to 8C are circuit diagrams showing other examples of the arrangements of the voltage generation circuits;

FIG. 9 is a graph showing  $V_g$ -I<sub>cell</sub> characteristics of the four-level flash memory when the voltage generation circuits are arranged as shown in FIGS. 8A to 8C;

FIG. 10 is a diagram showing a correlation between a reference current and a gate voltage in each of operating modes when the voltage generation circuits are arranged as shown in FIGS. 8A to 8C;

FIG. 11 is a circuit diagram showing a main part (a sense amplifier and a reference current generation circuit) of a four-level flash memory according to a second embodiment of the present invention;

FIGS. 12A and 12B are charts of operation waveforms obtained when the sense amplifier and reference voltage generation circuit are configured as shown in FIG. 11;

FIG. 13 is a circuit diagram showing another example of the arrangement of the reference current generation circuit shown in FIG. 11;

FIG. 14 is a circuit diagram showing a main part (a sense amplifier and a reference current generation



circuit) of a four-level flash memory according to a third embodiment of the present invention;

FIGS. 15A and 15B are charts of operation waveforms obtained when the sense amplifier and reference voltage generation circuit are configured as shown in FIG. 14;

FIG. 16 is a circuit diagram showing an arrangement of the main part of a binary flash memory in order to describe the prior art and its problems;

FIG. 17 is a graph showing  $V_g$ -I<sub>cell</sub> characteristics of the binary flash memory shown in FIG. 16;

FIG. 18 is a diagram showing a correlation between gate voltages in each of operating modes in the binary flash memory shown in FIG. 16; and

FIG. 19 is a diagram showing an example of the cell distribution with respect to a gate voltage in the binary flash memory shown in FIG. 16.

#### DETAILED DESCRIPTION OF THE INVENTION

Embodiments of the present invention will now be described with reference to the drawings.

(First Embodiment)

FIG. 1 shows an example of an arrangement of a four-level flash memory according to a first embodiment of the present invention. In FIG. 1, an address latch 11 latches an address from an address bus. An address counter 12 counts the address latched by the address latch 11. An address buffer 13 receives a count output

from the address counter 12 and outputs an internal address corresponding to a main cell (a selected memory cell to be processed) for reading, writing or erasing data. An IO buffer 14 controls the transmission and  
5 reception of read data and program data between a data bus and a data latch 15.

A memory cell array 21 has a plurality of main cells (a plurality of nonvolatile memory cells). The memory cell array 21 also has word, bit and source  
10 lines connected to the main cells. A sense amplifier 31 serving as a read circuit receives data of the main cells in the memory cell array 21 via a column gate circuit 32 in read mode. The sense amplifier 31 senses the received data and supplies it to the data latch 15.  
15 The sense amplifier 31 compares the current of the main cells flowing through a bit line by applying a read voltage to a word line and the current flowing through the reference cell that is provided in a reference current generation circuit 14 by applying a reference  
20 voltage to a reference word line. Data is therefore read out of the main cells. The configurations of the sense amplifier 31 and reference current generation circuit 41 will be described in detail later.

A write (program) circuit 33 receives program data  
25 from the data latch 15 in program mode and applies a program voltage to a corresponding bit line in the memory cell array 21 through the column gate circuit

32. In this case, the program circuit 33 can write cell data by applying a voltage to one or some of a word line, a bit line, a source line and a semiconductor region in which memory cells are formed.

5           A command register 34 holds various commands such as a program command and an erase command supplied from a control bus. A controller 35 receives the commands from the command register 34 and generates various control signals for controlling the respective circuits  
10 in the memory. A row decoder 36 receives an internal address from the address buffer 13 and selects a corresponding word line in the memory cell array 21. A column decoder 37 receives an internal address from the address buffer 13 and selectively drives a column  
15 gate in the column gate circuit 32 in response to the internal address. Thus, the selected bit line in the memory cell array 21 is connected to the sense amplifier 31 through the column gate circuit 32.

20           A charge pump circuit 38 serving as a step-up circuit steps up an external power supply voltage and generates a program voltage (e.g., 5V) and an erase voltage (e.g., 10V and -7V). The 5V program voltage generated from the charge pump circuit 38 is applied to the program circuit 33. The -7V erase voltage is  
25 applied to the memory cell array 21. The 10V erase voltage (Vpp) is applied to a regulator circuit 39. The regulator circuit 39 is configured by, e.g., a DA

converter circuit. The regulator circuit 39 generates various voltages  $V_{g\text{-hontai}}$ , which are applied to the word lines in program and read modes, from the voltage  $V_{pp}$  generated from the charge pump circuit 38. The  
5 voltages  $V_{g\text{-hontai}}$  generated from the regulator circuit 39 are applied to the word lines in the memory cell array 21, or the control gates in the main cells, through the row decoder 36.

The four-level flash memory includes an erase  
10 circuit (not shown) for erasing data of the main cells by applying a voltage to one or some of a word line, a bit line, a source line and a semiconductor region in which memory cells are formed.

The four-level flash memory also includes an erase  
15 verify circuit (not shown). In the present embodiment, the erase verify circuit applies a given voltage (e.g.,  $V_r$ ) to the word line of a main cell for erasure and the word line of the reference cell. The erase verify circuit compares the cell current of the selected main  
20 cell and that of the reference cell to detect whether data erasure is completed or not.

The four-level flash memory also includes a  
program verify circuit (not shown). In the present  
embodiment, the program verify circuit applies a given  
25 voltage (e.g.,  $V_r(PV1)$ ,  $V_r(PV2)$  and  $V_{sw2}(PV3)$ ) to the word line of a main cell for programming and applies a given voltage ( $V_r(PV1, PV2, PV3)$ ) to the word line

of the reference cell. The program verify circuit compares the cell current of the selected main cell and that of the reference cell to detect whether data programming is completed or not.

5           The above reference current generation circuit 41 has a plurality of reference cells and a voltage generation circuit (amplifier circuit) for amplifying the current flowing through the reference cells.

FIG. 2 shows a section of one of nonvolatile  
10       memory cells that make up the memory cell array 21 shown in FIG. 1. Referring to FIG. 2, an N-type well region (N-well) 21b is formed in a P-type semiconductor substrate (P-substrate) 21a. A P-type well region (P-well) 21c is formed in the N-type well region 21b.  
15       In the P-type well region 21c, an  $n^+$ -type source region 21d and an  $n^+$ -type drain region 21e are formed separately from each other. A floating gate 21f is formed on a channel region between the source and drain regions 21d and 21e with an insulation film 21-1  
20       interposed between the floating gate 21f and the channel region. A control gate 21g is formed on the floating gate 21f with an insulation film 21-2 interposed therebetween.

          A  $p^+$ -type contact region 21h is formed in a  
25       surface area of the P-type semiconductor substrate 21a. An  $n^+$ -type contact region 21i is formed in a surface area of the N-type well region 21b. A  $p^+$ -type contact

region 21j is formed in a surface area of the P-type well region 21c.

When a main cell is operated, a gate voltage  $V_{g\text{-hontai}}$  is applied to the control gate 21g. A drain voltage  $V_d$  is applied to the drain region 21e and a source voltage  $V_s$  is applied to the source region 21d. Furthermore, the same voltage as the source voltage  $V_s$  is applied to the contact regions 21i and 21j. A ground potential of 0V is applied to the contact region 21h.

The main cell stores "3" level data, "2" level data, "1" level data and "0" level data in accordance with the number of electrons stored in the floating gate 21f. In this case, the floating gate 21f differs from the control gate 21g in threshold voltage in accordance with the level of cell data to be stored.

The memory cell array 21 is made up of a plurality of memory cells so configured.

FIG. 3 shows a configuration of the main part of the four-level flash memory (e.g., a NOR type). Referring to FIG. 3, the memory cell array 21 includes a plurality of memory cells (main cells) MC arranged in matrix. The control gates (21g) of a plurality of main cells MC arranged in the same row are connected to their corresponding one of a plurality of word lines  $WL_0$  to  $WL_n$ . The drain regions (21e) of a plurality of main cells MC arranged in the same column are connected

to their corresponding one of a plurality of bit lines  
BL0 to BLk. In general, the memory cell array 21 is  
divided into a plurality of blocks. The source regions  
(21d) of a plurality of main cells MC in the same block  
5 are connected to their corresponding one of a plurality  
of source lines (not shown). The bit lines BL0 to BLk  
are connected to the sense amplifier 31 through their  
corresponding one of a plurality of select transistors  
ST0 to STk. The gates of the select transistors ST0  
10 to STk are connected to their corresponding one of  
a plurality of column lines COL0 to COLm.

The sense amplifier 31 includes three differential  
amplifiers (first, second and third sense amplifiers)  
31-1, 31-2 and 31-3 for comparing the cell current  
15 Icell of the main cells MC and three reference currents  
Iref0, Iref1 and Iref2. The sense amplifier 31  
generates a signal (2 bits) corresponding to cell data  
by a logical operation of outputs D0, D1 and D2 of the  
differential amplifiers 31-1, 31-2 and 31-3.

20 In the first embodiment, the cell current Icell of  
the main cells MC is supplied to the inverted input  
terminals of the differential amplifiers 31-1, 31-2  
and 31-3. The supply of the cell current Icell is  
performed through a p-type MOS transistor 31a and  
25 a bias transistor 31b that is configured by an n-type  
MOS transistor. In other words, the inverted input  
terminal of each of the differential amplifiers 31-1,

31-2 and 31-3 is connected to both the gate electrode and the drain region of the p-type MOS transistor 31a. A power supply Vcc is connected to the source region of the p-type MOS transistor 31a. The drain region of the p-type MOS transistor 31a is connected to that of the bias transistor 31b. A bias power line BIAS is connected to the gate electrode of the bias transistor 31b. The drain regions of the select transistors ST0 to STk, which are connected to the bit lines BL0 to BLk in the cell array 21, are connected to the source region of the bias transistor 31b.

On the other hand, the reference currents Iref0, Iref1 and Iref2 are supplied to the non-inverted input terminals of the differential amplifiers 31-1, 31-2 and 31-3 through p-type MOS transistors 31-1a, 31-2a and 31-3a and bias transistors 31-1b, 31-2b and 31-3b of n-type MOS transistors, respectively. In other words, the non-inverted input terminal of the differential amplifier 31-1 is connected to the gate electrode and drain region of the p-type MOS transistor 31-1a. A power supply Vcc is connected to the source region of the p-type MOS transistor 31-1a. The drain region of the p-type MOS transistor 31-1a is connected to the drain region of the bias transistor 31-1b. A bias power line BIAS is connected to the gate electrode of the bias transistor 31-1b. Similarly, the non-inverted input terminal of the differential amplifier 31-2 is



connected to the gate electrode and drain region of the  
p-type MOS transistor 31-2a. A power supply Vcc is  
connected to the source region of the p-type MOS  
transistor 31-2a. The drain region of the p-type MOS  
transistor 31-2a is connected to the drain region of  
5 the bias transistor 31-2b. A bias power line BIAS is  
connected to the gate electrode of the bias transistor  
31-2b. Similarly, the non-inverted input terminal of  
the differential amplifier 31-3 is connected to the  
10 gate electrode and drain region of the p-type MOS  
transistor 31-3a. A power supply Vcc is connected to  
the source region of the p-type MOS transistor 31-3a.  
The drain region of the p-type MOS transistor 31-3a is  
connected to the drain region of the bias transistor  
15 31-3b. A bias power line BIAS is connected to the gate  
electrode of the bias transistor 31-3b.

The reference currents Iref0, Iref1 and Iref2 are  
supplied to the source regions of the bias transistors  
31-1b, 31-2b and 31-3b. In other words, the reference  
20 current generation circuit 41 is connected to the bias  
transistors 31-1b, 31-2b and 31-3b to generate the  
reference currents Iref0, Iref1 and Iref2.

The reference current generation circuit 41  
includes generation circuits 41a, 41b and 41c for  
25 generating reference currents Iref0, Iref1 and Iref2,  
respectively. The generation circuits 41a, 41b and  
41c have reference column transistors 41a-1, 41b-1 and

41c<sub>-1</sub>, reference transistors QNR0, QNR1 and QNR2 and voltage generation circuits 41a<sub>-3</sub>, 41b<sub>-3</sub> and 41c<sub>-3</sub>. The reference column transistors 41a<sub>-1</sub>, 41b<sub>-1</sub> and 41c<sub>-1</sub> are each configured by an n-type MOS transistor.

5 The drain regions of the reference column transistors 41a<sub>-1</sub>, 41b<sub>-1</sub> and 41c<sub>-1</sub> are connected to their respective source regions of the bias transistors 31<sub>-1b</sub>, 31<sub>-2b</sub> and 31<sub>-3b</sub> in the sense amplifier 31. A reference column line RCOL is connected to the gate  
10 electrode of each of the reference column transistors 41a<sub>-1</sub>, 41b<sub>-1</sub> and 41c<sub>-1</sub>. The reference transistors QNR0, QNR1 and QNR2 are each configured by an n-type MOS transistor. The drain regions of the reference transistors QNR0, QNR1 and QNR2 are connected to their  
15 respective source regions of the reference column transistors 41a<sub>-1</sub>, 41b<sub>-1</sub> and 41c<sub>-1</sub>. The voltage generation circuits 41a<sub>-3</sub>, 41b<sub>-3</sub> and 41c<sub>-3</sub> are connected to their respective gate electrodes (reference word lines RWL0, RWL1 and RWL2) of the  
20 reference transistors QNR0, QNR1 and QNR2. The source regions of the reference transistors QNR0, QNR1 and QNR2 are connected to a ground potential. The voltage generation circuits 41a<sub>-3</sub>, 41b<sub>-3</sub> and 41c<sub>-3</sub> generate voltages that are to be applied to the gate electrodes  
25 of the reference transistors QNR0, QNR1 and QNR2, respectively.

FIGS. 4A to 4C illustrate examples of arrangements

of the above-described voltage generation circuits 41a-3, 41b-3 and 41c-3. The arrangements of the voltage generation circuits 41a-3, 41b-3 and 41c-3 are the same and only the threshold voltages  $V_{t0}$ ,  $V_{t1}$  and  
5  $V_{t2}$  of the reference cells RC0, RC1 and RC2 are different.

In FIGS. 4A to 4C, WS0, WS1 and WS2 indicate transconductances of current non-converting p-type MOS transistors PQa0, PQa1 and PQa2. WV0, WV1 and WV2  
10 indicate transconductances of current non-converting p-type MOS transistors PQb0, PQb1 and PQb2 in verify mode ( $V_{FYB} = L$ ). WR0, WR1 and WR2 indicate transconductances of current non-converting p-type MOS transistors PQc0, PQc1 and PQc2 in read mode ( $RDB = L$ ).  
15 The transconductances WS0, WS1 and WS2 are supplied to the source electrodes of the p-type MOS transistors PQa0, PQa1 and PQa2. The drain regions of the reference cells RC0, RC1 and RC2 are connected to the gate and drain regions of the p-type MOS transistors  
20 PQa0, PQa1 and PQa2 through bias transistors NQa of n-type MOS transistors, respectively. A gate voltage  $V_{read}$  that is equal to a read voltage is applied to the control gate of each of the reference cells RC0, RC1 and RC2. The source regions of the reference cells  
25 RC0, RC1 and RC2 are connected to a ground potential.

In the first embodiment, the threshold voltage of the reference cell (first reference cell) RC0 is  $V_{t0}$ ,

that of the reference cell (second reference cell) RC1 is  $V_{t1}$  ( $V_{t1} > V_{t0}$ ), and that of the reference cell RC2 (third reference cell) is  $V_{t2}$  ( $V_{t2} > V_{t1} > V_{t0}$ ).

5 The gates of the p-type MOS transistors PQb0, PQb1 and PQb2 and PQc0, PQc1 and PQc2 are connected to their corresponding gate and drain regions of the p-type MOS transistors PQa0, PQa1 and PQa2. The transconductances WV0, WV1 and WV2 are supplied to the source regions of the p-type MOS transistors PQb0, PQb1 and PQb2.

10 The drain regions of the p-type MOS transistors PQb0, PQb1 and PQb2 are connected to their respective source regions of p-type MOS transistors PQd. The gates of the p-type MOS transistors PQd are connected to their respective verify signal lines VFYB. The transconductances WR0, WR1 and WR2 are supplied to the source regions of the p-type MOS transistors PQc0, PQc1 and PQc2. The drain regions of the p-type MOS transistors PQc0, PQc1 and PQc2 are connected to the source regions of the p-type MOS transistors PQe. The gates of the p-type MOS transistors PQe are connected to their

20 corresponding read signal lines RDB. The drain regions of the p-type MOS transistors PQd and PQe are connected to each other and connected to the drain regions of bias transistors NQb of n-type MOS transistors and

25 the gates of n-type MOS transistors NQc, respectively.

The source regions of the bias transistors NQb are connected to their respective drain regions of the

n-type MOS transistors  $NQ_c$ . The source regions of the n-type MOS transistors  $NQ_c$  are each connected to a ground potential and the gates thereof are connected to their respective gates of the reference transistors  $QNR0$ ,  $QNR1$  and  $QNR2$ .

With the above circuit arrangement, the voltage generation circuits  $41a\_3$ ,  $41b\_3$  and  $41c\_3$  generate outputs corresponding to the threshold voltages  $V_{t0}$ ,  $V_{t1}$  and  $V_{t2}$  of the reference cells  $RC0$ ,  $RC1$  and  $RC2$  in data read and verify modes. The outputs of the voltage generation circuits  $41a\_3$ ,  $41b\_3$  and  $41c\_3$  are used to cause reference currents  $I_{ref0}$ ,  $I_{ref1}$  and  $I_{ref2}$  to flow through the reference transistors  $QNR0$ ,  $QNR1$  and  $QNR2$  and supplied to the gate electrodes (reference word lines  $RWL0$ ,  $RWL1$  and  $RWL2$ ) of the reference transistors  $QNR0$ ,  $QNR1$  and  $QNR2$ . In other words, the reference currents  $I_{ref0}$ ,  $I_{ref1}$  and  $I_{ref2}$  in read mode are generated when the p-type MOS transistor  $PQ_e$  turns on and the p-type MOS transistor  $PQ_d$  turns off to amplify the cell currents  $I_{refcell0}$ ,  $I_{refcell1}$  and  $I_{refcell2}$  in accordance with the ratio of transconductances  $WS0$ ,  $WS1$  and  $WS2$  to transconductances  $WR0$ ,  $WR1$  and  $WR2$ . The reference currents  $I_{ref0}$ ,  $I_{ref1}$  and  $I_{ref2}$  in verify mode are generated when the p-type MOS transistor  $PQ_d$  turns on and the p-type MOS transistor  $PQ_e$  turns off to amplify the cell currents  $I_{refcell0}$ ,  $I_{refcell1}$  and  $I_{refcell2}$  in accordance with the ratio of

transconductances WS0, WS1 and WS2 to transconductances WV0, WV1 and WV2.

Assume that the threshold voltage of a reference cell RC<sub>i</sub> is V<sub>t\_i</sub>, the cell current at the time of application of gate voltage V<sub>r</sub> is I<sub>refcell\_i</sub>, the ratio of WS<sub>i</sub> to WR<sub>i</sub> is 1 : a<sub>i</sub>, and the ratio of WS<sub>i</sub> to WV<sub>i</sub> is 1 : b<sub>i</sub>. The following equations are thus obtained:

$$\begin{aligned} I_{ref\_i} &= a_i I_{refcell\_i} \quad (\text{read mode}) \\ &= b_i I_{refcell\_i} \quad (\text{verify mode}) \quad \dots (1) \end{aligned}$$

If, therefore,  $b_0/a_0 < b_1/a_1 < b_2/a_2$  when  $V_{t0} < V_{t1} < V_{t2}$ , combinations of a<sub>i</sub> and b<sub>i</sub> are provided such that the verify margins I<sub>ref\_i\_v</sub> (v indicates a verify mode) and I<sub>ref\_i\_r</sub> (r indicates a read mode) are almost equal to each other, as shown in FIG. 5.

According to the first embodiment, the reference current generation circuit 41 includes at least a reference cell RC0 and a voltage generation circuit 41a<sub>-3</sub> for amplifying the cell current I<sub>refcell0</sub> flowing through the reference cell RC0. The reference current generation circuit 41 also includes at least a reference cell RC1 whose threshold voltage V<sub>t1</sub> is higher than the threshold voltage V<sub>t0</sub> of the reference cell RC0 and a voltage generation circuit 41b<sub>-3</sub> for amplifying the cell current I<sub>refcell1</sub> flowing through the reference cell RC1. The reference current generation circuit 41 also includes at least a

reference cell RC2 whose threshold voltage  $V_{t2}$  is higher than the threshold voltages  $V_{t0}$  and  $V_{t1}$  of the reference cells RC0 and RC1 and a voltage generation circuit 41c\_3 for amplifying the cell current  $I_{refcell2}$  flowing through the reference cell RC2. The first current amplification ratio, which is the ratio of the amplification factor ( $b_i$ ) of current in program verify mode to the amplification factor ( $a_i$ ) of current in data read mode in the voltage generation circuit 41a\_3, is set larger than "1". The second current amplification ratio, which is the ratio of the amplification factor of current in program verify mode to the amplification factor of current in data read mode in the voltage generation circuit 41b\_3, is set larger than "1". The third current amplification ratio, which is the ratio of the amplification factor of current in program verify mode to the amplification factor of current in data read mode in the voltage generation circuit 41c\_3, is set larger than "1".

FIG. 6 shows an example of the cell distribution with respect to the reference current  $I_{ref}$ . In FIG. 6, v indicates the verify mode and r indicates the read mode. In the present embodiment, the lower limit of the cell distribution corresponding to the state "0" is compensated with reference current  $I_{ref0v}$ . Similarly, the lower limit of the cell distribution corresponding to the state "1" is compensated with reference current

Ireflv. Further, the lower limit of the cell distribution corresponding to the state "2" is compensated with reference current Iref2v. Reference current Iref0r is located between the cell distribution corresponding to the state "0" and the cell distribution corresponding to the state "1". Reference current Ireflr is located between the cell distribution corresponding to the state "1" and the cell distribution corresponding to the state "2". Reference current Iref2r is located between the cell distribution corresponding to the state "2" and the cell distribution corresponding to the state "3".

FIG. 7 shows a correlation among reference current Iref, gate voltage Vg-hontai of main cell MC, and gate voltage Vg-ref of reference cell RC in each operating mode. For example, reference current Iref2r or Iref2v has only to be used as the reference current Iref in over-erase verify (OEV) mode and program verify (PV3) mode for defining the data program state "2".

In data read mode (Read mode), the reference current Iref is Iref0r, Ireflr and Iref2r, and the gate voltage Vg-hontai of main cell MC and the gate voltage Vg-ref of reference RC are each Vr (= 5.5V). In erase verify mode (EV mode), the reference current Iref is Iref0v, and the gate voltage Vg-hontai of main cell MC and the gate voltage Vg-ref of reference RC are each Vr. In over-erase verify mode (OEV mode), the



reference current Iref is Iref2r (or Iref2v), the gate voltage Vg-hontai of main cell MC is Vsw1 (= 2V), and the gate voltage Vg-ref of reference RC is Vr. In program verify mode (PV1 mode) for defining data  
5 program state "0", the reference current Iref is Iref1v, and the gate voltage Vg-hontai of main cell MC and the gate voltage Vg-ref of reference RC are each Vr. In program verify mode (PV2 mode) for defining data program state "1", the reference current Iref is  
10 Iref2v, and the gate voltage Vg-hontai of main cell MC and the gate voltage Vg-ref of reference RC are each Vr. In program verify mode (PV3 mode) for defining data program state "2", the reference current Iref is Iref2r (or Iref2v), the gate voltage Vg-hontai of main  
15 cell MC is Vsw2 (= 6.5V) and the gate voltage Vg-ref of reference RC is Vr.

For the sake of simplification, it is desirable to use "1" for all amplification factors a<sub>i</sub>; however, of course, other values can be used. When all the  
20 amplification factors are "1" and Iref<sub>i\_v</sub> - Iref<sub>i\_r</sub> is equal to IO, the verify margins of the respective distributions can be fixed if the following equation is set based on the characteristics diagram of FIG. 6:

$$(b_i - 1)gm(V_r - V_{t_i}) = IO$$

25 or

$$b_i = 1 + IO/gm/(V_r - V_{t_i}) \dots (2)$$

where gm is transconductance of cell current

Irefcell<sub>i</sub>.

If the threshold voltage  $V_{t_i}$  of reference cell RC<sub>i</sub> is adjusted in both program and erase modes, the verify margin can be secured by the above equation (2).

5       As described above, the reference current Iref<sub>r</sub> in read mode and the reference current Iref<sub>v</sub> in verify mode are generated from the same reference cell RC<sub>i</sub> based on the conversion rate (current ratio) according to the level. Thus, the verify margin can stably be  
10       secured. In the verify margin of cells of different distributions, the conversion rate is set for each of the reference cells RC<sub>i</sub>. A fixed verify margin can thus be secured for the cells of any distribution.

      In particular, two reference currents can be  
15       generated from one reference cell. Consequently, the number of reference cells can be reduced, which is very useful for downsizing the memory.

      In the embodiment described above, the present invention is applied to a four-level flash memory. The  
20       invention is not limited to this but can be applied to a binary flash memory. When the invention is applied to a binary flash memory, a circuit for generating an erase verify voltage  $V_{ev}$  becomes unnecessary and thus the binary flash memory can be decreased in size.

25       In the foregoing embodiment, the reference current generation circuit includes three reference cells (see FIGS. 4A to 4C). The present invention is not limited

to this. The reference current generation circuit can be made up of, e.g., two reference cells. FIGS. 8A to 8C show an example of the arrangement of a voltage generation circuit including two reference cells in a four-level flash memory. In this example, a reference cell is omitted from a voltage generation circuit for generating a voltage that is to be applied to the gate electrode (RWL2) of the reference transistor QNR2.

In the above example, for instance, a voltage generation circuit 41c\_3' generates reference current Iref2 on the basis of the bias voltage PBIAS of the voltage generation circuit 41b\_3'. Thus, the number of reference cells RC\_i is reduced and the arrangement of the circuit is simplified.

In the above example, when  $V_{t0} < V_{t1}$ , parameters are determined to satisfy the following expressions:  $b_0/a_0 < b_1/a_1 < b_2/a_2$ ,  $b_1 > b_2$ , and  $a_1 > a_2$ . In other words, the first current amplification ratio, which is the ratio of the amplification factor ( $b_i$ ) of current in program verify mode to the amplification factor ( $a_i$ ) of current in read mode in the voltage generation circuit 41a\_3 is set larger than "1". The second current amplification ratio, which is the ratio of the amplification factor of current in program verify mode to the amplification factor of current in read mode in the voltage generation circuits 41b\_3 and 41c\_3' is set larger than "1" and the first current amplification

ratio is smaller than the second current amplification ratio.

FIG. 9 shows  $V_g$ -I<sub>cell</sub> characteristics in the foregoing circuit arrangement. FIG. 10 shows a relationship among reference current I<sub>ref</sub>, gate voltage  $V_{g\text{-hontai}}$  and gate voltage  $V_{g\text{-ref}}$  in each operating mode.

(Second Embodiment)

FIG. 11 shows a main part (a sense amplifier and a reference current generation circuit) of a four-level flash memory according to a second embodiment of the present invention. In the second embodiment, the sense amplifier includes two differential amplifiers.

As shown in FIG. 11, a sense amplifier 31' includes two differential amplifiers 31-1 and 31-2, a latch circuit 51a, an inverter circuit 51b and two AND circuits 51c and 51d. The inverted input terminals of the differential amplifiers 31-1 and 31-2 are connected to a power supply V<sub>cc</sub> through a resistance element R<sub>ra</sub> and supplied with a cell current I<sub>cell</sub> of a main cell MC. The noninverted input terminal of the differential amplifier 31-1 is connected to a reference current generation circuit 41' and supplied with a reference current I<sub>ref1</sub> from a generation circuit (I<sub>ref</sub> conversion circuit) 41b'. The noninverted input terminal of the differential amplifier 31-2 is connected to the reference current generation circuit

41' and supplied with a reference current  $I_{ref0}$  from a generation circuit ( $I_{ref}$  conversion circuit) 41a' or a reference current  $I_{ref2}$  from a generation circuit ( $I_{ref}$  conversion circuit) 41c'.

5           The latch circuit 51a latches an output signal D0 of the differential amplifier 31<sub>-1</sub> and supplies the latched signal to the inverter circuit 51b and AND circuit 51c. The AND circuit 51c generates a reference select signal RCOL0 from both the latched signal and  
10           a control signal 2ndcol. The reference select signal RCOL0 is supplied to the gate of a reference column transistor 41a<sub>-1</sub> of the generation circuit 41a'.  
The AND circuit 51d generates a reference select signal RCOL2 from both an inverted signal, which is obtained  
15           by inverting the latched signal output from the latch circuit 51a by the inverter circuit 51b, and a control signal 2ndcol. The reference select signal RCOL2 is supplied to the gate of a reference column transistor 41c<sub>-1</sub> of the generation circuit 41c'.

20           In the second embodiment, the generation circuits 41a', 41b' and 41c' have substantially the same arrangement. The generation circuits 41a', 41b' and 41c' include their respective voltage generation circuits 41a<sub>-3</sub>, 41b<sub>-3</sub> and 41c<sub>-3</sub> (see, for example,  
25           FIGS. 4A to 4C) that are so configured that only the threshold voltages  $V_{t0}$ ,  $V_{t1}$  and  $V_{t2}$  of the reference cells RC0, RC1 and RC2 are different. The generation

circuits also each include a resistance element Rrb connected to the power supply Vcc.

Of the reference select signals RCOL0, RCOL1 and RCOL2, the reference select signal RCOL1 is always  
5 activated (high level) and supplied to the gate of a reference column transistor 41b<sub>-1</sub> of the generation circuit 41b'. One of the reference select signals RCOL0 and RCOL2 is activated in accordance with the logic of the output D0 of the differential amplifier  
10 31<sub>-1</sub>. The output D0 of the differential amplifier 31<sub>-1</sub> indicates higher two ones or lower two ones of four threshold voltage levels. The output D1 of the differential amplifier 31<sub>-2</sub> indicates which one of the higher two threshold voltage levels is higher/lower or  
15 indicates which one of the lower two threshold voltage levels is higher/lower.

The above circuit arrangement allows the number of differential amplifiers to reduce and thus the area of the circuit can be decreased more than that in the first embodiment. Moreover, the output D0 of the  
20 differential amplifier 31<sub>-1</sub> and the output D1 of the differential amplifier 31<sub>-2</sub> are assigned to I/O terminals of different addresses and the outputs D0 and D1 are supplied in this order. With these specifica-  
25 tions, access time can be prevented from being lengthened greatly.

FIGS. 12A and 12B show operation waveforms of the

four-level flash memory shown in FIG. 11. FIG. 12A shows the operation waveforms obtained when data is "11" (indicated by a solid line) and data is "10" (indicated by a broken line). FIG. 12B shows the operation waveforms obtained when data is "01" (indicated by a solid line) and data is "00" (indicated by a broken line). In FIG. 12A and FIG. 12B, WL/RWL represents selected word line WL and reference cell word line RWL.

As shown in FIG. 12A, the potential of the word lines WL and RWL of the selected memory cell and reference cell and reference select signal RCOL1 become a high level at time T0. Thus, the differential amplifier 31-1 compares the cell current  $I_{cell}$  of the selected main cell (hereinafter referred to as the selected cell) MC with the reference current  $I_{ref1}$  generated from the generation circuit 41b'. When the cell current  $I_{cell}$  of the selected cell MC is larger than the reference current  $I_{ref1}$ , the output D0 of the differential amplifier 31-1 becomes high (time T1). Thus, the output (RCOL0) of the AND circuit 51c becomes high (time T2) in synchronization with the control signal 2ndcol.

In contrast, as shown in FIG. 12B, when the cell current  $I_{cell}$  of the selected cell MC is smaller than the reference current  $I_{ref1}$ , the output D1 of the differential amplifier 31-2 becomes low (time T1).

Thus, the output (RCOL2) of the AND circuit 51d becomes high (time T2) in synchronization with the control signal 2ndcol.

5 In the foregoing cases, the outputs D0 and D1 of the differential amplifiers 31<sub>-1</sub> and 31<sub>-2</sub> indicate one of four states "0", "1", "10" and "11" in accordance with the level of each of the outputs. In other words, the output D0 of the differential amplifier 31<sub>-1</sub> is supplied to the outside of the sense amplifier 31' when  
10 it is determined. On the other hand, the differential amplifier 31<sub>-2</sub> compares the cell current I<sub>cell</sub> of the selected cell MC with the reference current I<sub>ref0</sub> or I<sub>ref2</sub>. Subsequent to the output D0, the output D1 of the differential amplifier 31<sub>-2</sub> is supplied to the  
15 outside when it is determined. Thus, the four-level flash memory performs a data read operation.

In a program verify operation performed when program data is "00", the cell current I<sub>cell</sub> of the main cell MC is compared with the current generated by  
20 amplifying the reference current I<sub>ref2</sub>. Then, when the cell current I<sub>cell</sub> becomes smaller than the reference current I<sub>ref2</sub>, additional program data is changed to "11". In a program verify operation performed when program data is "01", the cell current I<sub>cell</sub> of the  
25 main cell MC is compared with the current generated by amplifying the reference current I<sub>ref1</sub>. Then, when the cell current I<sub>cell</sub> becomes smaller than the reference



current Iref1, additional program data is changed to "11". In a program verify operation performed when program data is "10", the cell current Icell of the main cell MC is compared with the current generated  
5 by amplifying the reference current Iref0. Then, when the cell current Icell becomes smaller than the reference current Iref0, additional program data is changed to "11".

By changing the amplification factors of the  
10 currents in read mode and verify mode as described above, a verify margin can be secured.

FIG. 13 shows another example of the arrangement of the reference current generation circuit (Iref conversion circuit) shown in FIG. 11. In FIG. 13,  
15 an erase verify control circuit for an erase verify operation is added to the generation circuit 41b' shown in FIG. 11.

Referring to FIG. 13, a generation circuit 41b'' includes an erase verify control circuit 61. The  
20 circuit 61 decreases a reference current in erase verify mode more than that in normal data read mode and has three n-type MOS transistors NQ1, NQ2 and NQ3. More specifically, the drain region of the n-type MOS transistor NQ1 is connected to the drain region  
25 of a p-type MOS transistor PQd, the drain region of a p-type MOS transistor PQe, the drain region of a bias transistor NQb, the gate of an n-type MOS transistor

NQc, and the gate of the n-type MOS transistor NQ1.

A bias power line BIAS is connected to the gate of the n-type MOS transistor NQ1 and the drain region of the n-type MOS transistor NQ2 is connected to the source

5 region thereof. An EV mode signal line is connected to the gate of the n-type MOS transistor NQ2 and the drain region of the n-type MOS transistor NQ3 is connected to the source region thereof. The source region of the n-type MOS transistor NQ3 is connected to a ground  
10 potential.

In the above case, the size of each of the transistors is set in such a manner that the reference current has a minimum value in erase verify mode (EV = High, RDB = Low), a maximum value in program verify  
15 mode, and an intermediate value between them in data read mode. This arrangement allows the reference current to decrease in erase verify mode. The current conversion ratio in the generation circuit is parameter essential to the verify margin. It is thus desirable  
20 that all the transistors be of the same size and the current conversion ratio be adjusted according to the number of transistors of the same gate length and gate width.

(Third Embodiment)

25 FIG. 14 a main part (a sense amplifier and a reference current generation circuit) of a four-level flash memory according to a third embodiment of the

present invention. In the third embodiment, the sense amplifier includes one differential amplifier.

As shown in FIG. 14, a sense amplifier 31'' includes a differential amplifier 31-1, latch circuits 51a and 51e, an inverter circuit 51b and two AND circuits 51c and 51d. The inverted input terminal of the differential amplifier 31-1 is connected to a power supply Vcc through a resistance element Rra and supplied with a cell current Icell of a main cell MC. The noninverted input terminal of the differential amplifier 31-1 is connected to a reference current generation circuit 41'' and supplied with one of reference currents Iref1, Iref0 and Iref2 generated from generation circuits (Iref conversion circuits) 41b', 41a' and 41c'.

The latch circuit 51a latches an output signal of the differential amplifier 31-1, which serves as an output Dout of the sense amplifier 31'', in response to a latch control signal Lat0, and outputs the latched signal as an output D0. The latched signal of the latch circuit 51a is supplied to the inverter circuit 51b and AND circuit 51c. The AND circuit 51c generates a reference select signal RCOL0 from both the latched signal of the latch circuit 51a and a control signal 2ndcol. The reference select signal RCOL0 is supplied to the gate of a reference column transistor 41a-1 of the generation circuit 41a'. The AND circuit 51d

generates a reference select signal RCOL2 from both an inverted signal, which is obtained by inverting the latched signal output from the latch circuit 51a by the inverter circuit 51b, and a control signal 2ndcol.

5 The reference select signal RCOL2 is supplied to the gate of a reference column transistor 41c<sub>-1</sub> of the generation circuit 41c'. The latch circuit 51e latches an output signal (Dout) of the differential amplifier 31<sub>-1</sub> in response to a latch control signal Lat1 and  
10 outputs the latched signal as an output D1.

Of the reference select signals RCOL0, RCOL1 and RCOL2, the reference select signal RCOL1 is always activated (high level) and supplied to the gate of a reference column transistor 41b<sub>-1</sub> of the generation  
15 circuit 41b' in accordance with the supply of a control signal 1stcol to a buffer circuit 52. One of the reference select signals RCOL0 and RCOL2 is activated in accordance with the logic of the output D0 of the latch circuit 51a. The output D0 of the latch circuit  
20 51a indicates higher two ones or lower two ones of four threshold voltage levels. The output D1 of the latch circuit 51e indicates which one of the higher two threshold voltage levels is higher/lower or indicates which one of the lower two threshold voltage levels is  
25 higher/lower.

The above circuit arrangement allows the number of differential amplifiers to reduce further and thus the

area of the circuit can be decreased more advantageously than that in the second embodiment. Moreover, the output D0 of the latch circuit 51a and the output D1 of the latch circuit 51e are assigned to I/O terminals of different addresses and the outputs D0 and D1 are supplied in this order. With these specifications, access time can be prevented from being lengthened greatly.

FIGS. 15A and 15B show operation waveforms of the four-level flash memory shown in FIG. 14. FIG. 15A shows the operation waveforms obtained when data is "11" (indicated by a solid line) and data is "10" (indicated by a broken line). FIG. 15B shows the operation waveforms obtained when data is "01" (indicated by a solid line) and data is "00" (indicated by a broken line). In FIG. 15A and FIG. 15B, WL/RWL represents selected word line WL and reference cell word line RWL.

As shown in FIG. 15A, the potential of the word lines WL and RWL of the selected memory cell and reference cell and reference select signal RCOL1 become a high level at time T0. Thus, the differential amplifier 31-1 compares the cell current I<sub>cell</sub> of the selected main cell (hereinafter referred to as the selected cell) MC with the reference current I<sub>ref1</sub> generated from the generation circuit 41b'. When the cell current I<sub>cell</sub> of the selected cell MC is larger

than the reference current  $I_{ref1}$ , the output of the differential amplifier 31<sub>-1</sub> becomes high. The output of the differential amplifier 31<sub>-1</sub> is latched by the latch circuit 51a at the falling edge of the latch control signal Lat0 (immediately before time T2).

Thus, the output (RCOL0) of the AND circuit 51c becomes high in synchronization with the control signal 2ndcol. Thus, the output of the differential amplifier 31<sub>-1</sub> is latched by the latch circuit 51e at the falling edge of the latch control signal Lat1 (time T3).

In contrast, as shown in FIG. 15B, when the cell current  $I_{cell}$  of the selected cell MC is smaller than the reference current  $I_{ref1}$ , the output of the differential amplifier 31<sub>-1</sub> becomes low. The output of the differential amplifier 31<sub>-1</sub> is latched by the latch circuit 51a at the falling edge of the latch control signal Lat0 (immediately before time T2). Thus, the output (RCOL2) of the AND circuit 51d becomes high in synchronization with the control signal 2ndcol. Thus, the output of the differential amplifier 31<sub>-1</sub> is latched by the latch circuit 51e at the falling edge of the latch control signal Lat1 (time T3).

In the above circuit arrangement, too, the four-level flash memory can perform the data read, program verify and erase verify operations as in the second embodiment.

The foregoing embodiments are all directed to

a four-level flash memory. The present invention is not limited to such a flash memory. For example, the invention can be applied to a binary flash memory.

When it is applied to a binary flash memory, no  
5 circuits for generating an erase verify voltage  $V_{ev}$  are required and accordingly the binary flash memory can be decreased in size.

Additional advantages and modifications will readily occur to those skilled in the art. Therefore,  
10 the invention in its broader aspects is not limited to the specific details and representative embodiments shown and described herein. Accordingly, various modifications may be made without departing from the spirit or scope of the general inventive concept as  
15 defined by the appended claims and their equivalents.